2010

Optimal design for generalized linear models with a multinomial response

Gwenda Thompson

University of Wollongong

Optimal Design for Generalized Linear Models with a Multinomial Response

A thesis submitted in fulfillment of the requirements for the award of the degree

Doctor of Philosophy

from

University of Wollongong

by

Gwenda Thompson BMath (Hons) UOW

School of Mathematics and Applied Statistics

May 2010

I, Gwenda Piera Thompson, declare that this thesis, submitted in fulfillment of the requirements for the award of Doctor of Philosophy in the School of Mathematics and Applied Statistics, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Gwenda Thompson

May 2010

Acknowledgment

This thesis would not have been possible without the support of my family, friends and colleagues.

I wish to express my sincere gratitude to Associate Professor Ken Russell for providing this opportunity and being a patient and supportive supervisor. A special thank you to Dr Damian Collins for all his help and friendship. Thanks to Professors David Steel and David Griffiths for help and advice and to the staff of the School of Mathematics and Applied Statistics.

Funding for my scholarship has been provided by the Australian Research Council and grants were awarded by the Statistical Society of Australia, the Centre for Statistical and Survey Methodology, Design and Analysis of Experiments Conference and Industry and Investment NSW.

I wish to thank my parents, family and most importantly, my husband, Mark.

Abstract

The objective of optimal experimental design is to determine the values of the predictor variables at which to take observations in order to maximize the quality of data obtained by the experiment. When the response variable is, at least approximately, independent and normally distributed with a constant variance, an optimal design can be obtained by accessing the prolific body of established classical design techniques. When these assumptions are not satisfied, the design solution is not easily obtainable.

The difficulty in designing an experiment for a response with a nonlinear model is that the optimal design depends upon the values of the unknown parameters of the model, but the model parameters are estimated from the results obtained after running the experiment. This has been a major factor contributing to the complexity, and hindering the development, of design methodology for nonlinear models. As a consequence, the majority of results, including those in this thesis, are obtained utilizing numerical techniques.

We examine two design methods for a multinomial response, utilizing the generalized linear model framework. One technique, D-optimality, is based on minimizing the volume of the confidence ellipsoid of the parameter estimates and the other, IMSE-optimality, is based on minimizing the mean squared error of the estimated probabilities of occurrence of the possible responses. These methods are developed using point estimates of the parameters of the linear

predictor. When estimates are unavailable, we use partial knowledge about the parameters to make designs more robust to uncertainty in the parameter values. We also develop designs which allow for some uncertainty in the model by examining different types of predictors.

The experimental design space is shown to play an important role, especially in IMSE-optimality, and a quantitative method is presented to determine the design space based on the limits of the underlying multinomial probabilities.

Although the two design methods examined are based on different criteria, our simulations strongly suggest that they are asymptotically equivalent. The D-optimal design algorithm is shown to be faster to implement, compared to the IMSE-optimal design algorithm, and is free to determine the number of support points. The IMSE-optimal approach requires more computation as it requires taking expected values of estimated probabilities and specifying the number of observations at each support point.

These design techniques are applied to a practical example, a dose response experiment, and we develop design methods allowing for some uncertainty in the form of the predictor and the values of the parameter estimates.

Contents

List of Figures

List of Tables